Context Matters: How an Ecological-Belonging Intervention Can Reduce Inequities in STEM

SARAH P. HAMMARLUND, CHERYL SCOTT, KEVIN R. BINNING, AND SEHOYA COTNER

Doubts about belonging in the classroom are often shouldered disproportionately by students from historically marginalized groups, which can lead to underperformance. Ecological-belonging interventions use a classroom-based activity to instill norms that adversity is normal, temporary, and surmountable. Building on prior studies, we sought to identify the conditions under which such interventions are effective. In a chemistry course (study 1), students from underrepresented ethnic backgrounds underperformed relative to their peers in the absence of the intervention. This performance gap was eliminated by the intervention. In an introductory biology course (study 2), there were no large performance gaps in the absence of the intervention, and the intervention had no effect. Study 2 also explored the role of the instructor that delivers the intervention. The intervention boosted scores in the classrooms of instructors with a fixed (versus growth-oriented) intelligence mindset. Our results suggest that ecological-belonging interventions are more effective in more threatening classroom contexts.

Keywords: education, STEM equity, sense of belonging, belonging intervention

University-level science, technology, engineering, and mathematics (STEM) courses often fail to provide an equitable experience for all students, leading to the exclusion of students from historically underserved groups (Olsen and Rordan 2012, National Research Council 2018). Students from underrepresented racial and ethnic backgrounds, women, and first-generation college students often perform worse than their peers in STEM courses, even when analyses correct for differences in precourse preparedness (Lohfink and Paulsen 2005, Hyde et al. 2008, Matz et al. 2017, Salehi et al. 2019, Salehi et al. 2020). Among other factors, classroom social contexts—the daily interactions that students have with their peers, teaching assistants (TAs), and instructors—may contribute to disparities. For example, by virtue of their exposure to American society, most students and instructors in US classrooms are aware of the stereotypes that pertain to particular groups in academic settings (Steele 1997). When stereotypes are in the air, they may create psychological threat among negatively stereotyped students (e.g., stereotype threat; Steele and Aronson 1995, Steele 1997), which can impair their learning (Taylor and Walton 2011) and hinder their high-stakes test performance (Schamder and Sedikides 2018, Salehi et al. 2019). Importantly, nonstereotyped group members also commonly experience threat responses during interactions with negatively stereotyped others, particularly when they first meet (e.g., Blascovich et al. 2001, Goff et al. 2008). Perhaps unsurprisingly, students from negatively stereotyped groups tend to report lower levels of confidence and feel less like they belong in science than do their nonstereotyped counterparts (Walton and Cohen 2007, Rainey et al. 2018, Gopalan and Brady 2020). Negative affective experiences can have an adverse impact on participation, performance, and persistence (Cohen and Garcia 2008, Schmader and Sedikides 2018). In particular, a reduced sense of belonging can affect how students interpret and respond to adversity (Wheeler and Petty 2001, Cohen and Garcia 2008). For example, if a student receives a low exam score, they may interpret the score as confirmation that people like them cannot succeed in the course and do not belong in STEM. This can lead to a self-fulfilling cycle of low performance and a reduced sense of belonging that may result in withdrawal from STEM (Cohen and Garcia 2008, Rainey et al. 2019).

Researchers have recently attempted to negate the power of stereotypes in the classroom by changing classroom norms about the meaning and implications of adversity (e.g., when students struggle to learn difficult concepts or receive a poor exam grade). One recent study employed an ecological-belonging intervention that was designed to provide an alternative social norm within the classroom, instilling the idea that all students experience adversity and that challenges are normal, temporary, and surmountable (Binning

BioScience XX: 1–10.
https://doi.org/10.1093/biosci/biab146

et al. 2020). The approach was developed from prior social-belonging interventions (Walton and Cohen 2007, 2011, Walton et al. 2015, Yeager et al. 2016, Murphy et al. 2020), with the alteration that it is implemented in the classroom with discussion between instructors and peers in an effort to change the classroom ecology and intersubjective norms.

The roughly 30-minute intervention exercise consists of reflective writing about the challenges that students anticipate encountering in the course, reading testimonials from more advanced students (from both stereotyped and nonstereotyped backgrounds) who overcame past challenges, and a classroom discussion with peers and the instructor that was designed to reinforce and establish the intervention message as a local norm. Classroom discussions may be especially powerful for creating a common understanding that everyone struggles, regardless of their gender or ethnicity, which may rob factors such as stereotype threat of their power. Binning and colleagues (2020) found that an ecological-belonging intervention improved the performance of historically underperforming students. Specifically, they found improvement in course grades for ethnic minorities in an introductory biology course and for women in an introductory physics course.

Although these findings are promising, there are well-known concerns about the replicability of psychological science (Open Science Collaboration 2015) and social science more broadly (Camerer et al. 2018). Many high-profile studies have failed to be replicated when studied in new contexts. Social–psychological interventions in education are no exception; this work has yielded inconsistent replications (Walton 2014, Schwartz et al. 2016), and the evidence suggests that social context may powerfully moderate intervention effects (Walton 2014, Schwartz et al. 2016, Binning and Browman 2020, Walton and Yeager 2020).

In the present research, we sought to identify the contexts in which an ecological-belonging intervention is most effective at promoting equity in college STEM classrooms. Across two studies, we trained current instructors and TAs to lead the intervention, most of whom taught two sections of the same course. Each instructor was then assigned to conduct the intervention in one of their class sections and to conduct their other section as usual with no changes. This design allowed for a strong test of the intervention's robustness and its potential for scalability beyond where it was first developed. It also allowed us to examine which features of the instructors' course context are predictive of where the intervention may be effective. First, the results of Binning and colleagues (2020) suggest that the ecological-belonging intervention is most effective for students who experience underperformance: The intervention was effective where underperformance occurred (e.g., women in physics) but not where it did not (e.g., women in biology). Second, a contextual factor that remains underexplored is instructors' beliefs about students' intelligence. Many characteristics of instructors are important for student learning (e.g., warmth and approachability; Widmeyer and Loy 1988, Rainey et al. 2019). One characteristic that has recently been shown to affect students' feelings of threat and performance is the instructor's mindset (Canning et al. 2019, 2021, Muenks et al. 2020). Instructors with so-called fixed mindsets believe that students are born with a certain amount of intelligence and cannot do anything to change their abilities. In contrast, instructors with growth mindsets believe that intelligence is like a muscle that grows stronger with practice (Dweck 2006, Dweck and Yeager 2019). Emerging research has shown that instructors' mindsets not only affect student outcomes (Muenks et al. 2020) but may also moderate the effectiveness of sociocognitive educational interventions (Yeager et al. 2021). Instructor mindset may therefore influence the potential for ecological-belonging interventions to be effective. However, the way in which instructors' mindsets may moderate the ecological-belonging intervention is unclear. On one hand, having an instructor with a growth-oriented mindset may help to reinforce the message of the belonging intervention, which may lead to greater positive effects of the intervention in classrooms with growth-oriented instructors, similar to the findings of Yeager and colleagues (2021). On the other hand, students of instructors with more fixed mindsets have been shown to experience greater psychological vulnerability and less belonging (Muenks et al. 2020). When an instructor with a fixed mindset conducts an ecological-belonging intervention at the start of the term, this may create a positive first impression that prevents those harmful effects of a fixed-mindset instructor (a so-called halo effect; Asch 1946, Nisbett and Wilson 1977). We would therefore see greater positive effects of the intervention in classrooms with fixed-mindset instructors.

We report the results of ecological-belonging interventions conducted in two undergraduate STEM courses at a large public research university. Study 1 was conducted in an introductory chemistry lecture course for nonchemistry major students, and study 2 was performed in the laboratory sections associated with a nonmajors introductory biology course. These courses had different structures and different performance gaps between students from various demographic groups. We sought to identify the contexts in which the intervention is and is not effective at eliminating performance gaps between marginalized and nonmarginalized students. On the basis of the results of Binning and colleagues (2020), we hypothesized that the intervention would only improve the scores of students from demographic groups that showed performance gaps in control sections that were taught as usual. We also hypothesized that the effect of the ecological-belonging intervention would be moderated by instructor mindset.

Methods

We conducted two studies of ecological-belonging interventions.

Study 1: Introductory chemistry. Study 1 involved a nonmajors' lecture-based introductory chemistry course taught by one instructor with two sections during the same term. Each section had approximately 300 students. One section was randomly chosen to receive the intervention, and the other (control, or business-as-usual) section was taught without any changes.

The ecological-belonging intervention took place during two separate class periods spaced 1 week apart at the beginning of the Fall 2019 semester. During the first intervention activity, the instructor asked their students to write a brief reflection about the challenges they anticipated facing during the course (see supplemental material section S1 for the intervention materials). The reflections were anonymous. The instructor then collected the written responses. During the second intervention activity, the instructor presented four quotes and told the students that they were from students who had taken the course in the past. These quotes were adapted from Binning and colleagues (2020) and originated from focus group interviews from students at another institution (see Walton and Brady 2020). We made minor changes to these quotes to suit our institution. The quotes were intended to represent the general concerns and challenges expressed by students. For example, a quote from Aniyah, a junior, read,

> “When I first got here, I was worried because it seemed like there weren’t many students like me. And I was really struggling with some of the chemistry concepts. It felt like everyone else was doing just fine, but I just wasn’t sure if I was cut out for the course. At some point during the first semester, I came to realize that, actually, a lot of other students were struggling, too. And I started to look at struggling as a positive thing. After I struggled with a hard problem and then I talked to other classmates and my TA about the solution, I realized that all that effort was worth it because it helped me learn and remember much more.”

Next, the instructor asked the students to discuss these statements in small groups using three prompts:

- *What are some common themes across several of the quotes we read?*
- *Why do you think that sometimes students don’t realize that other people are also struggling with the course?*
- *Why and how does people’s experience change over time? What do people do that helps them improve their experience with time?*

Finally, the instructor called on groups to share some of their responses to the above prompts with the class, and the instructor facilitated a brief whole-class discussion.

A total of 610 students participated in study 1. The control section had $n = 271$ students, and the intervention section had $n = 339$ students. However, because of missing data (see below) and students who withdrew at the start of the course, we analyzed data from $n = 247$ students in the control section and $n = 303$ students in the intervention section.

We examined four demographic variables: the students’ gender, their college generation, whether they belong to a minoritized and underrepresented ethnic or racial group, and whether they are of Asian descent. We use *male* and *female* to describe gender, but we recognize that these refer to biological sex rather than gender and may not represent how students identify. We used institutional data that unfortunately only included binary options. The total sample was 58% female and 42% male. We categorized students as first generation if no parent or guardian has received a 4-year undergraduate degree from a college or university. The total sample was 18.5% first generation and 81.5% continuing generation. We categorized students from minoritized and underrepresented ethnic or racial groups as PEER (for *persons excluded because of their ethnicity or race*; Asai 2020). PEER status is based on institutional ethnicity data and includes students whose ethnicity is listed as American Indian, Black, Hawaiian, and Hispanic students. This demographic grouping is often referred to as URM (for *underrepresented minority*; National Center for Science and Engineering Statistics 2019). Asian and White students, who are overrepresented in STEM relative to the general population, are designated as non-PEER students. We note that categorizing students as PEER or non-PEER is imperfect. Some individuals may not identify with this term, and any aggregation hides differences in the experiences of students of different ethnicities within the groups (Bhatti 2021). The total sample was 1.1% American Indian, 22% Asian, 5.5% Black, 0.33% Hawaiian, 2.8% Hispanic, 65% White, and 2.3% unknown. For the final demographic variable, we separated Asian students for analysis because, although Asian students are represented in STEM at similar rates as in the general population (National Center for Science and Engineering Statistics 2019), they may face unique challenges that White students do not face. In other words, we used two dummy codes to represent ethnicity: The PEER variable separates non-White and non-Asian students from White and Asian students, and the Asian variable separates Asian students from non-Asian students. White students are therefore the contrast variable for the analysis. All four demographic categories were equally represented in the control section and the intervention section (gender, $\chi^2(1) = 3.73$, $p = .053$; college generation, $\chi^2(1) = 1.45$, $p = .230$; PEER status, $\chi^2(1) = 0.608$, $p = .436$; Asian status, $\chi^2(1) = 2.88e{-30}$, $p = 1$).

We used grand mean-centered ACT score and high school GPA as metrics of precourse preparedness. There was no difference between the students in the intervention and control sections in ACT score and high school GPA (p » .05). A minority of students (111 students) were missing either their ACT score or their high school GPA. We assigned those students the mean scores to avoid excluding them from the study. We also performed our analyses with these students excluded. The average total course score did not differ between the students missing either their ACT or high school GPA data and the students with no missing data (two-sample t-test: $t(123) = 0.54$, $p = .59$).

The dependent variable was total course score, which is the number of course points obtained out of 100 possible percentage points at the end of the semester. A passing score in the course was 46 out of 100 points, corresponding to a C– letter grade.

We obtained demographic data from the university registrar's student database and the course scores from the instructor. The individuals were anonymized by a researcher who was not associated with the study. In total, we excluded 75 students from the analysis because of missing data. The majority of these students (67 students) were excluded because they dropped or withdrew from the course and were therefore missing total score data. The remaining eight students were excluded because of missing demographic data. The number of students that dropped or withdrew from the course did not differ between the control and the intervention sections ($\chi^2(1) = 0.39$, $p = .53$). Access to institutional data and grades was considered exempt from full review by the university's Institutional Review Board (STUDY00000800). All of the students gave informed consent to participate in this research in a survey emailed to them at the start of the semester.

We first analyzed whether any demographic variables had an effect on the students' final course scores in the control section with multiple regression analysis. Next, using the Aiken and West (1991) procedure for testing for statistical interactions, we investigated whether the intervention had an effect on final course scores and whether there were any interactions between demographic variables and the intervention. We included ACT score and high school GPA as variables in the multiple regression models to control for the students' precourse preparedness. We used R version 3.6.0 for the study 1 analyses.

Study 2: Introductory biology. Study 2 involved a nonmajors introductory biology course that largely serves students in the prehealth sciences and the natural sciences beyond biology. The course has both lecture and lab components. Three lecture sections run concurrently each semester, and each lecture section is associated with between 9 and 14 lab sections, with maximum 24 students per section. The intervention took place in the lab sections during the first 2 weeks of the Fall 2019 semester. The lab sections (32 total) were led by graduate and undergraduate TAs. A total of 16 TAs taught the lab sections. Of those, 12 led two lab sections. For those TAs, one of their sections was randomly chosen to receive the intervention and the other was a control (business-as-usual) section. One TA led three lab sections; two of them were randomly assigned to receive the intervention, and the third was assigned to be a control section. Two TAs taught only one section; both received the intervention. The remaining TA taught one intervention section and two sections that were unknown; we excluded those two sections. In total, we therefore analyzed data from 30 of the 32 sections. The intervention activities were identical to the procedure described above for study 1.

A total of 588 students participated in study 2; $n = 324$ students received the intervention, and $n = 264$ students were in the control sections. As in study 1, we used four demographic variables: gender (36% male, 64% female), college generation (76% continuing generation, 24% first generation), PEER status (American Indian, Black, Hawaiian, and Hispanic; 80% non-PEER, 16% PEER, 4% unknown), and Asian status (80% non-Asian, 16% Asian, 4% unknown). Three demographic categories were equally represented in the control section and in the intervention section (gender, $\chi^2(1) = 2.89$, $p = .089$; college generation, $\chi^2(1)= 0.110$, $p = .740$; Asian status, $\chi^2(1) = 0.068$, $p = .790$). PEER students were overrepresented in the control sections ($\chi^2(1) = 13.4$, $p = .00026$).

The 16 TAs completed a survey that included two Likert-scale questions about their mindset about intelligence used by Canning and colleagues (2019): "Consider the undergraduate students you will teach and respond to this quote: 'To be honest, students have a certain amount of intelligence, and they really can't do much to change it,'" and "Consider the undergraduate students you will teach and respond to this quote: 'Your intelligence is something about you that you can't change very much.'" The TAs selected a response from a 6-point Likert scale (1, *agree*, 2, *somewhat agree*, 3, *neither agree nor disagree*, 4, *somewhat disagree*, 5, *disagree*, and 6, *strongly disagree*). We averaged the responses to the two questions. The TAs' mindset scores ranged from 2 to 6, with a mean of 4.98 and a standard deviation of 1.16. This variable was then mean-centered.

As in study 1, we used the students' grand mean-centered ACT score and high school GPA as a measure of precourse preparedness. There was no difference between students in the intervention and control sections for ACT score and high school GPA (p » .05). A minority of the students (118 students) were missing either ACT score or high school GPA. We assigned those students the mean scores in order to avoid excluding them from the study. We also performed our analyses with these students excluded. The average total course score did not differ between the students missing either their ACT or high school GPA data and the students with no missing data (two-sample t-test, $t(151) = -1.9$, $p = .06$).

As in study 1, the dependent variable was total course score, which is the number of course points obtained out of 100 possible percentage points. The total score includes points earned in the lab portion of the course in addition to points earned in the lecture portion (most of the students earned all of the possible lab points). A passing score in the course was 56 out of 100 percentage points, which corresponded to a D, although the students taking the course pass/fail needed 60 or more points to pass the course.

We obtained demographic data from the university registrar's student database and course scores from the instructors. The individuals were anonymized by a researcher who was not associated with the study. In the analyses, we used restricted maximum likelihood estimation to handle missing data by using all available data (Raudenbush and Bryk 2002).

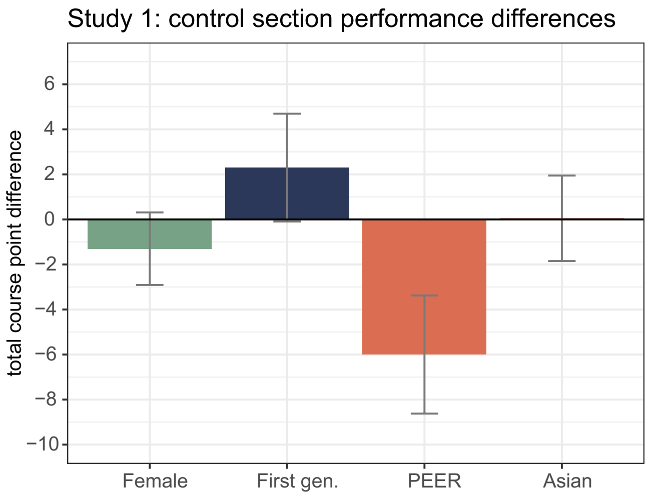

Figure 1. PEER students are underserved in introductory chemistry. Analysis of total course points (out of 100) within the control (business-as-usual) section ($n = 247$). PEER indicates underrepresented racial or ethnic minority students, which encompasses American Indian, Hawaiian, Black and Hispanic students. The control section included 146 female students, 39 first-generation students, 28 PEER students, and 57 Asian students. The* y*-axis value of 0 represents the performance of the reference group for each demographic group. For the female students, the reference group is male students. For the first-generation students, it is continuing-generation students. For the PEER students, it is White and Asian students. For the Asian students, the reference group is students of all other ethnicities. The PEER students performed around six percentage points lower than White and Asian students did ($t(234) = -2.3$, $p = .022$). The error bars represent the standard error.

However, runtime deletion excluded 12 students who were missing their total course score data, bringing the analyzed model to $n = 576$. Access to institutional data and grades was considered exempt from full review by the university's Institutional Review Board (STUDY00000800). All of the students and TAs gave informed consent to participate in this research in a survey emailed to them at the start of the semester.

To account for the nestedness of the data, we employed a three-level hierarchical multiple regression model (Raudenbush and Bryk 2002) using HLM 7.0 software in study 2. The students were divided into different lab sections, some sharing TAs, and some sharing lecture sections. The TAs ($n = 16$) were modeled at level 3, each TA's sections ($n = 30$) were modeled at level 2, and the students were modeled at level 1 ($n = 588$). The dependent variable, total course points, was, in turn, partially nested within one of three different lecture sections. As such, we entered two dummy codes at level 1 to account for the differences in average performance across the three lecture sections. Also at level 1, we controlled for precourse preparation by entering the grand mean-centered ACT score and the grand mean-centered high school GPA. We also included dummy-coded variables to capture gender, generation, PEER, and Asian demographic categories. At level 2, we included the classroom condition code (0, control; 1, treatment). In addition, to explore the potential contribution of the TAs' mindsets to their students' performance, at level 3, we included the TAs' mean-centered mindset score.

The data and analysis scripts for both study 1 and study 2 are available in an online repository (Hammarlund et al. 2021).

Results

We studied the effects of ecological-belonging interventions in two introductory STEM courses.

Study 1: Introductory chemistry. In order to understand whether certain students were underserved by the course, we examined the scores in the control section to identify performance gaps. We identified demographic groups whose members underperformed relative to our expectations based on the precourse preparedness metrics (high school GPA and ACT score). Within the control section, the underrepresented ethnic and racial minority (PEER) students scored 6.0 percentage points lower than non-PEER (White and Asian) students ($n = 247$, $b = -6.04$, standard error [SE] = 2.62, $p = .022$, $t(234) = -2.30$). We found no performance differences within other demographic categories (gender, college generation, and Asian status; figure 1, supplemental material sections S2.1 and S2.2).

Because we controlled for precourse preparedness in this analysis, the results indicate that underperformance emerged in the course among the PEER students, above and beyond prior differences in the students' precourse preparedness. Although there are many reasons performance gaps may emerge, we suggest that one factor may be that stereotypes were "in the air" (Steele 1997), which contributed to underperformance among the PEER students. We hypothesized that the ecological-belonging intervention would reduce this performance gap.

We found no main effect of the intervention on total course points ($n = 550$, $b = 0.610$, SE = 1.09, $t(527) = 0.559$, $p = .576$; supplemental material section S2.3), indicating that the intervention had no effect on overall student performance. However, we did find an interaction between PEER status and the intervention, where the PEER students who received the intervention showed increased performance (figure 2, supplemental material section S2.4). Among the non-PEER (White and Asian) students, the intervention had no effect ($b = -0.189$, SE = 1.15, $t(526) = -0.165$, $p = .869$), but among the PEER students, the intervention was associated with an increase of 7.5 points ($b = 7.54$, SE = 3.36, $t(526) = 2.24$, $p = .025$), erasing the performance gap between the White and Asian students and the PEER students. For many of the students, an increase of 7.5 points meant the difference between passing and failing the course.

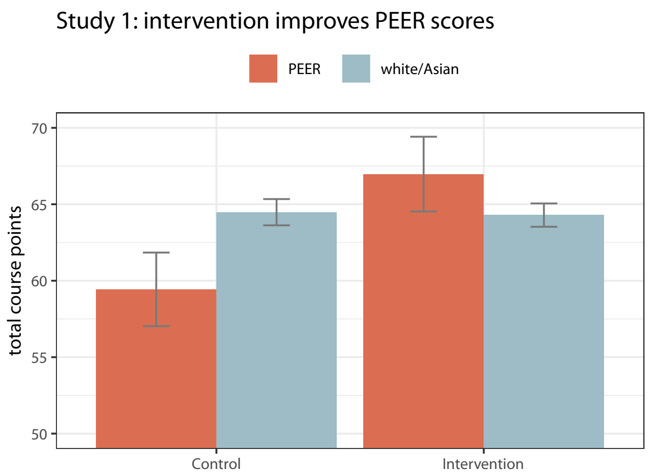

Figure 2. PEER students benefit from the ecological-belonging intervention in introductory chemistry. The analysis shows an interaction effect between PEER status and the intervention. The orange bars show the PEER (American Indian, Hawaiian, Black and Hispanic) students, and the gray bars show the non-PEER (White and Asian) students. When the PEER students received the intervention, their total course scores increased by 7.54 percentage points (t(526) = 2.242, p = .025), erasing the performance gap between the PEER and the non-PEER students. The control section contained 247 students total, with 28 PEER students, and the intervention section contained 303 students total, with 27 PEER students. The error bars represent the standard error.

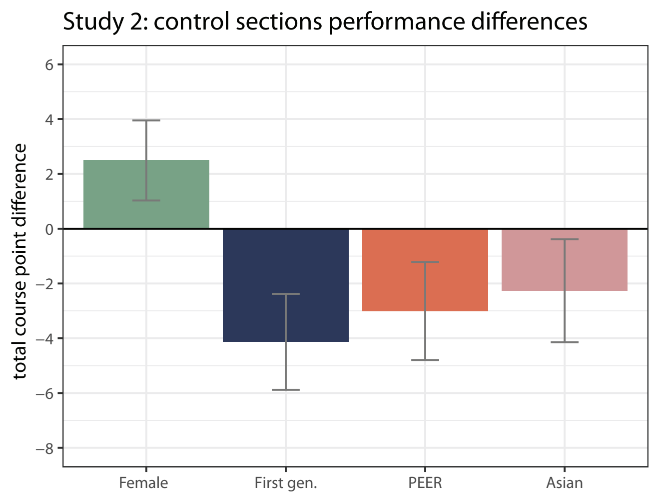

Figure 3. First-generation college students are underserved in introductory biology. An analysis of the total course points (out of 100) of the students within the control (business-as-usual) sections (n = 264). PEER indicates underrepresented racial or ethnic minority students, which encompasses American Indian, Hawaiian, Black, and Hispanic students. The control sections included 179 female students, 66 first-generation students, 58 PEER students, and 45 Asian students. The y-axis value of 0 represents the performance of the reference group for each demographic group. For the female students, the reference group is male students. For the first-generation students, it is continuing-generation students. For the PEER students, it is White and Asian students. For Asian students, the reference group is students of all other ethnicities. The first-generation college students performed around four points lower than the continuing-generation students (t(251) = 1.75, p = .019). No other demographic categories showed significant performance differences. The error bars represent the standard error.

There were no interactions between the intervention and the other demographic categories. These patterns were also seen in the analyses with casewise deletion of the students with missing ACT scores and high school GPA data (supplemental material section S2.5).

Study 2: Introductory biology. As in study 1, we first examined the scores in the control sections to identify performance gaps. Within the control lab sections, the first-generation college students scored 4.13 points lower than their continuing-generation peers ($t(251) = -2.36$, $p = .019$; figure 3, supplemental material section S3.1). We found no performance differences within the other demographic categories (gender, PEER, and Asian status). We hypothesized that first-generation students' scores may improve as a result of the belonging intervention.

The test of the study hypotheses proceeded in several steps; in the first analyses, we attempted to directly replicate the effect seen in our study 1 and in study 1 of Binning and colleagues (2020), where the belonging intervention specifically improved the scores of demographic groups that showed a performance gap. Our analyses revealed that there was no overall main effect of the intervention on the students' performance at level 2 ($b = 0.20$, SE = 0.93, $t(28) = 0.22$, $p = .830$; supplemental material section S3.2), nor was there a differential cross-level interaction between the intervention and first-generation status ($b = -1.05$, SE = 2.14, $t(564) = -0.49$, $p = .623$; supplemental material section S3.3). This means that the first-generation college students' scores did not improve as a result of the intervention. Furthermore, there was no cross-level interaction between the intervention and PEER status ($b = -2.28$, SE = −2.57, $t(564) = -0.89$, $p = .376$).

To explore the effect of the TAs' intelligence mindset on the students' performance and to examine whether the intervention had effects among the TAs with a low (versus high) growth mindset scores, we tested for a TA mindset score (level 3) × intervention (level 2) interaction on the students' course grades. The analysis indicated a significant interaction ($b = -1.44$, SE = 0.36, $t(28) = 4.01$, $p = .001$; figure 4, supplemental material section S3.4). An analysis of the effect of the intervention across the levels of TA mindset revealed that the intervention had a

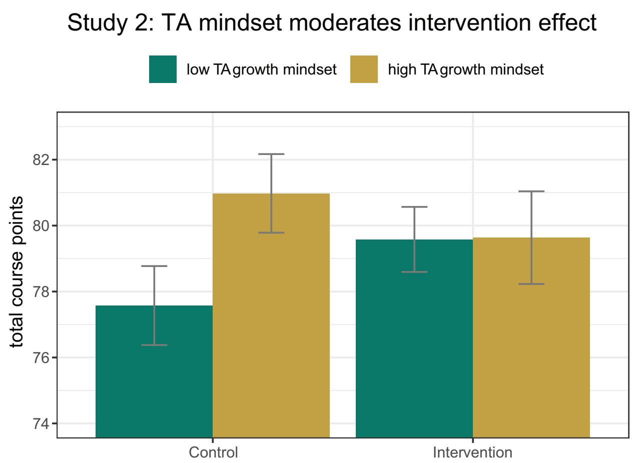

Figure 4. Teaching assistants' views of students' intelligence moderate the effects of the ecological-belonging intervention in introductory biology. TA mindset, a continuous variable, is shown as low (1 standard deviation below the mean, the green bars) and high (1 standard deviation above the mean, the gold bars). Within the control lab sections, the students with TAs with a more growth-oriented mindset performed better than the students of TAs with a more fixed mindset (t(14) = 3.02, p = .010). Within the intervention lab sections, TA mindset had no effect on student performance (t(14) = 0.06, p = .957). The students of TAs with more fixed mindsets (the green bars) performed better if they received the intervention (t(28) = 2.15, p = .04). Among the students of TAs with more growth-oriented mindsets (the gold bars), the intervention had no significant effect (t(28) = –1.34, p = .193). The error bars represent the standard error.

positive effect among the TAs with a relatively low growth mindset scores (1 standard deviation below the mean; $b = 2.01$, SE = 0.93, $t(28) = 2.15$, $p = .04$). By contrast, the intervention had no effect in the lab sections led by the TAs with a strong growth mindset (1 standard deviation above the mean; $b = -1.34$, SE = 1.00, $t(28) = -1.34$, $p = .193$).

Notably, in the control classrooms, the TAs' mindset was significantly correlated with their students' course scores ($b = 1.47$, SE = 0.49, $t(14) = 3.02$, $p = .010$). That is, the TAs' mindsets were predictive of their students' course grades: Within control classrooms, students of TAs with high growth mindsets performed better than students of TAs with relatively fixed mindsets about intelligence. However, within the classrooms in which the TAs delivered the belonging intervention, TA mindset had no effect on the students' scores ($b = 0.02$, SE = 0.41, $t(14) = 0.06$, $p = .957$). These results are robust to the exclusion of students with missing ACT scores and high school GPAs from the data set (supplemental material section S3.5). There was no three-way interaction among the intervention, TA mindset, and any demographic variable.

Discussion

We predicted that the intervention would help students in groups that underperformed relative to their peers in the control sections of the two courses. In other words, the intervention would only be effective for students who experienced a sense of threat and uncertainty about belonging that leads to a performance gap. This is consistent with findings from similar studies (Schwartz et al. 2016, Binning et al. 2020). In study 1, where the underrepresented ethnic and racial minority students (American Indian, Hawaiian, Black and Hispanic students, or PEER students) underperformed relative to the White and Asian students, we found that the intervention did improve the PEER students' scores. In study 2, the PEER students did not significantly underperform in the control condition. However, the first-generation college students underperformed relative to the continuing-generation students. We found that the intervention had no effect on either the PEER or the first-generation students.

In study 2, we also predicted that the instructors' mindsets about intelligence would affect the effectiveness of the intervention. This could happen in two ways: The intervention may be more effective in the classrooms of instructors with a growth mindset because that mindset could reinforce the message of the intervention, or the intervention could be more effective in the classrooms of fixed-mindset instructors because the intervention may compensate for the negative effects of having an instructor with a fixed mindset. Our data supported the second prediction. The results from the control sections confirm that having a fixed-mindset instructor can have negative effects; the students of instructors with more fixed mindsets had lower scores than the students of instructors with growth mindsets, replicating the main effect described by Canning and colleagues (2019). The intervention specifically improved the scores of students whose TAs had a fixed view of intelligence, but it had no effect on the students whose TAs had a growth mindset. When a TA with a fixed mindset performed the intervention, this may have changed their students' perceptions of the TA, perhaps causing the students to view their instructor more positively. This positive impression may have made the TAs' more fixed mindsets irrelevant. In addition, the intervention may have provided the students with tools to counter negative signals from fixed-mindset TAs. The students of TAs with high growth mindset scores may have already felt less threatened and, therefore, did not stand to benefit from the intervention.

Overall, our results support the idea that ecological-belonging interventions are most effective in threatening contexts. The intervention bolstered the scores of PEER students in contexts in which they were underperforming, but it had no effect in contexts where they were not. The intervention improved the performance of students whose TA had a fixed mindset, but it had no effect when the TAs had a growth mindset. In other words, the intervention appears to be effective only where it is needed. We also found no negative effects of the intervention, so, although the intervention does not improve performance across the board, it does not *harm* students' performance.

The results have promising practical implications. Importantly, this research demonstrates that the ecological-belonging intervention can be effectively exported to new university contexts, effectively taught to instructors, and effectively delivered with impact. Unlike those in Binning and colleagues (2020), the instructors who delivered the intervention in our studies were not discipline-based education researchers. In fact, many were undergraduate and graduate-level TAs with limited experience in the classroom and little—if any—pedagogical training. This intervention training was brief and would be easy to implement at a larger scale (adaptable materials are provided in supplemental material section S1). Indeed, we know of several instructors at other institutions around the United States who plan to implement similar interventions, enabling us to further contextualize the conditions in which a belonging exercise is helpful. In addition, this is the first time ecological-belonging interventions have been implemented in nonmajor courses and in courses taken mainly by 2nd- through 4th-year students. This suggests that these interventions have power beyond so-called gateway major courses taken at the start of students' university experiences. Finally, the intervention had a positive effect in both a large (approximately 300 students) lecture course and in lab sections of approximately 24 students, suggesting that the intervention's effect may be robust to changes in class sizes and contexts.

An important caveat is that, although a threatening classroom context may be necessary for the intervention to have an effect, it is not sufficient. The lack of intervention benefits for first-generation students in study 2 illustrates this point. There was a small but significant performance gap between the first-generation and the continuing-generation students in the control sections, but this gap was also present in the intervention sections. Previous research has shown that social–psychological interventions are capable of improving outcomes among first-generation college students (Harackiewicz et al. 2014, Stephens et al. 2014). Although it is difficult to speculate about the causes of null effects, we believe that one important distinction may be the visibility of the students' PEER status relative to their first-generation status. In particular, whereas PEER status is commonly (though not always) visible (e.g., due to physical characteristics), first-generation status is usually not visible. Therefore, when student engage in the intervention, they would not know whether a peer was a first-generation student unless that fellow student volunteered the information. If the intervention works by shaping intersubjective norms, perhaps a lack of visibility makes the intervention less effective. Indeed, research on so-called concealable stigmatized identities (Crocker et al. 1998, Quinn and Chaudoir 2009) and hidden identities (Henning et al. 2019) has demonstrated the importance of the visibility of the demographic feature, as well as the potential benefits of mutual disclosure of concealable identities (Cooper et al. 2020). Whether visibility affects the impact of the intervention is an important question for future research.

Several standard limitations of human-subjects research apply to our study. Specifically, the participating students and the TAs represented a range of backgrounds, experiences, and axes of diversity. We were unable to control for all sources of variety, some of which may have affected our findings. We were also unable to explore the students' intersecting identities, because of limited sample sizes. In addition, we only investigated one response variable (overall performance), so we do not know what, if any, additional impacts the intervention may have had. Critically, we did not measure students' actual sense of belonging in either population. Furthermore, study 1 had one intervention and one control section, limiting our ability to extrapolate beyond the study population. However, our findings resonate with those of prior work on belonging interventions (Binning et al. 2020) and instructor mindset (Canning et al. 2019, Muenks et al. 2020) while also adding nuance to our current understanding of the impact of course-level interventions.

Our findings suggest several important avenues for future research. We began with reference to the concerns about the replicability of psychological science, and our findings confirm that the results are dependent on the context in which the intervention is delivered (Van Bavel et al. 2016). Therefore, we reiterate prior calls for further replication in other contexts (i.e., beyond large research institutions, in other disciplines, with different documented challenges). Future work could illustrate whether the intervention has a lasting effect. Related interventions have shown effects that persist over multiple years and in varied social settings (e.g., after students graduate; Goyer et al. 2017, Brady et al. 2020), but the duration and generality of ecological-belonging interventions need further study. In addition, our findings about the interaction between instructor mindset and the intervention contrast with those of other studies about the contexts in which social–psychological interventions are most successful. Recent evidence has indicated that similar interventions are more effective in supportive contexts—for example, in contexts in which students are exposed to teachers with a growth mindset (Yeager et al. 2021) or in contexts that support challenge-seeking norms (Yeager et al. 2016). Our findings, in contrast, showed that a social–psychological intervention was more effective when the instructors had a more potentially threatening fixed mindset. We suspect that a key difference may be in who delivered the interventions. In contrast to these prior studies, the present intervention was delivered by the same instructors whose mindsets were studied. A long history of research on person–perception has shown that the characteristics people learn early about the targets of social judgment can shape and filter their subsequent judgments (Asch 1946, Nisbett and Wilson 1977). Therefore, given that the intervention was delivered as one of the first activities the students did in their class, perhaps leading the intervention created a sort of halo effect that changed how the students subsequently perceived their instructor. More research about the importance of the intervention leader's role in the course is warranted.

In addition to refining our contextual understanding, future work should target a mechanistic understanding: *Why* do these brief interventions work? Establishing mechanism will likely require a rigorous qualitative investigation; this could involve interviewing students at different timepoints to investigate how the intervention affected their sense of belonging (à la Rainey et al. 2018), confidence about overcoming adversity, awareness of other students' struggles, feelings about whether other students like them and care about them, and their perceptions of their instructors. Using instruments that measure belonging (such as the Psychological Sense of School Membership Scale; Goodenow 1993) and experience sampling methodology to measure real-time student experiences (Muenks et al. 2020) could also elucidate whether the mechanism by which we're assuming the intervention works is accurate (Schwartz et al. 2016).

We also suggest that instructor characteristics beyond mindset may affect the intervention's effectiveness. As more interventions are conducted and assessed, it will become possible to conduct a meta-analysis to explore the predictive power of other instructor characteristics on student outcomes. For example, does it matter if the instructor is the same race, ethnicity, or gender as the students experiencing a threat? Similarly, can an intervention compensate for having an instructor that does not match a student's identity?

Finally, while overseeing the interventions, we noted, anecdotally, that merely reading the concerns of their students (which we did not require) could affect TAs, leading them to express surprise at how worried the students were about the material, interacting with the instructor, or managing the course requirements. We are currently exploring the impact of the intervention on the TAs themselves, and we are motivated by one primary question: Can delivering an intervention foster greater instructor empathy for their students?

Our findings add to the current dialogue on both the promise and caveats of psychosocial interventions, specifically those designed to mitigate barriers to equity in STEM education. These interventions are not a panacea for all challenges to equity in STEM, but that does not invalidate their implementation. Rather, this work calls for continued documentation of the criteria under which belonging interventions are effective. Critically, any intervention should be envisioned as a short-term solution. Our collective goal as STEM educators should be to realize a future in which our scientists represent our current population, and belonging interventions are no longer necessary.

Acknowledgments

This project was supported by the SEISMIC collaboration (www.seismicproject.org). Funding for the SEISMIC project has been provided by the Alfred P. Sloan Foundation and the Participating Institutions. This project was also supported by the Biology Teaching Assistant Project (BioTAP), an NSF-Funded Research Coordination Network (grant no. DBI 1539903). SPH was supported by a NSF Graduate Research Fellowship. We thank members of the Cotner lab for feedback and support, especially Dr. Sadie Hebert for assistance with data collection and curation.

Supplemental material

Supplemental data are available at *BIOSCI* online.

References cited

Aiken LS, West SG. 1991. Multiple Regression: Testing and Interpreting Interactions. Sage.

Asai DJ. 2020. Race matters. Cell 181: 754–757.

Asch SE. 1946. Forming impressions of personality. Journal of Abnormal and Social Psychology 41: 258–290.

Bhatti HA. 2021. Toward "inclusifying" the underrepresented minority in STEM education research. Journal of Microbiology and Biology Education 22: e00202–e00221.

Binning KR, Browman AS. 2020. Theoretical, ethical, and policy implications for conducting social-psychological interventions to close educational achievement gaps. Social Issues and Policy Review 14: 182–216.

Binning KR, Kaufmann N, McGreevy EM, Fotuhi O, Chen S, Marshman E, Kalender ZY, Limeri L, Betancur L, Singh C. 2020. Changing social contexts to foster equity in college science courses: An ecological-belonging intervention. Psychological Science 31: 1059–1070.

Blascovich J, Mendes WB, Hunter SB, Lickel B, Kowai-Bell N. 2001. Perceiver threat in social interactions with stigmatized others. Journal of Personality and Social Psychology 80: 253–267.

Brady ST, Cohen GL, Jarvis SN, and Walton GM. 2020. A brief social-belonging intervention in college improves adult outcomes for black Americans. Science Advances 6: 1–12.

Camerer CF, et al. 2018. Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. Nature Human Behaviour 2: 637–644.

Canning EA, Muenks K, Green DJ, Murphy MC. 2019. STEM faculty who believe ability is fixed have larger racial achievement gaps and inspire less student motivation in their classes. Science Advances 5: 1–7.

Canning EA, Ozier E, Williams HE, AlRasheed, R, Murphy MC. 2021. Professors who signal a fixed mindset about ability undermine women's performance in STEM. Social Psychological and Personality Science. https://doi.org/10.1177/19485506211030398.

Cohen GL, Garcia J. 2008. Identity, belonging, and achievement: A model, interventions, implications. Current Directions in Psychological Science 17: 365–369.

Cooper KM, Gin LE, Brownell SE. 2020. Depression as a concealable stigmatized identity: What influences whether students conceal or reveal their depression in undergraduate research experiences? International Journal of STEM Education 7: 1–8.

Crocker J, Major B, Steele C. 1998. Social stigma. Pages 504–553 in Gilbert DT, Fiske ST, Lindzey G, eds. The Handbook of Social Psychology, vol. 2. McGraw-Hill.

Dweck CS. 2006. Mindset: The New Psychology of Success. Random House.

Dweck CS, Yeager DS. 2019. Mindsets: A view from two eras. Perspectives on Psychological Science 14: 481–496.

Goff PA, Steele CM, Davies PG. 2008. The space between us: Stereotype threat and distance in interracial contexts. Journal of Personality and Social Psychology 94: 91–107.

Goodenow C. 1993. The psychological sense of school membership among adolescents: Scale development and educational correlates. Psychology in the Schools 30: 79–90.

Gopalan M, Brady ST. 2020. College students' sense of belonging: A national perspective. Educational Researcher 49: 134–137.

Goyer JP, Garcia J, Purdie-Vaughns V, Binning KR, Cook JE, Reeves SJ, Apfel N, Taborsky-Barba S, Sherman DK, Cohen GL. 2017. Self-affirmation facilitates minority middle schoolers' progress along college trajectories. Proceedings of the National Academy of Sciences 114: 7594–7599.

Hammarlund SP, Scott C, Binning KR, Cotner S. 2021. Context Matters: How an "Ecological-Belonging" Intervention Can Reduce Inequities in STEM. Zenodo. https://doi.org/10.5281/zenodo.4891850.

Harackiewicz JM, Canning EA, Tibbetts Y, Giffen CJ, Blair SS, Rouse DI, Hyde JS. 2014. Closing the social class achievement gap for first-generation students in undergraduate biology. Journal of Educational Psychology 106: 375–389.

Henning JA, Ballen CJ, Molina SA, Cotner S. 2019. Hidden identities shape student perceptions of active learning environments. Frontiers in Education 4: 1–13.

Hyde JS, Lindberg SM, Linn MC, Ellis AB, Williams CC. 2008. Gender similarities characterize math performance. Science 321: 494–495.

Lohfink MM, Paulsen MB. 2005. Comparing the determinants of persistence for FGCS. Journal of College Student Development 46: 409–428.

Matz RL, Koester BP, Fiorini S, Grom G, Shepard L, Stangor CG, Weiner B, McKay TA. 2017. Patterns of gendered performance differences in large introductory courses at five research universities. AERA Open 3: 1–12.

Muenks K, Canning EA, LaCosse J, Green DJ, Zirkel S, Garcia JA, Murphy MC. 2020. Does my professor think my ability can change? Students' perceptions of their STEM professors' mindset beliefs predict their psychological vulnerability, engagement, and performance in class. Journal of Experimental Psychology: General 149: 2119–2144.

Murphy MC, Gopalan M, Carter ER, Emerson KT, Bottoms BL, Walton GM. 2020. A customized belonging intervention improves retention of socially disadvantaged students at a broad-access university. Science Advances 6: 1–7.

National Center for Science and Engineering Statistics. 2019. Women, Minorities, and Persons with Disabilities in Science and Engineering. National Science Foundation.

National Research Council. 2018. Indicators for Monitoring Undergraduate STEM Education. National Academies Press.

Nisbett RE, Wilson TD. 1977. The halo effect: Evidence for unconscious alteration of judgments. Journal of Personality and Social Psychology 35: 250–256.

Olsen S, Rordan DG. 2012. Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics. Executive Office of the President.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. Science 349: 1–8.

Rainey K, Dancy M, Mickelson R, Stearns E, Moller S. 2018. Race and gender differences in how sense of belonging influences decisions to major in STEM. International Journal of STEM Education 5: 1–14.

Rainey K, Dancy M, Mickelson R, Stearns E, Moller, S. 2019. A descriptive study of race and gender differences in how instructional style and perceived professor care influence decisions to major in STEM. International Journal of STEM Education 6: 1–13.

Raudenbush SW, Bryk AS. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods 2nd ed. Sage.

Quinn DM, Chaudoir SR. 2009. Living with a concealable stigmatized identity: The impact of anticipated stigma, centrality, salience, and cultural stigma on psychological distress and health. Journal of Personality and Social Psychology 97: 634–651.

Salehi S, Cotner S, Ballen CJ. 2020. Variation in incoming academic preparation: Consequences for minority and first-generation students. Frontiers in Education 5: 1–14.

Salehi S, et al. 2019. Gender performance gaps across different assessment methods and the underlying mechanisms: The case of incoming preparation and test anxiety. Frontiers in Education 4: 1–14.

Schmader T, Sedikides C. 2018. State authenticity as fit to environment: The implications of social identity for fit, authenticity, and self-segregation. Personality and Social Psychology Review 22: 228–259.

Schwartz DL, Cheng KM, Salehi S, Wieman C. 2016. The half empty question for socio-cognitive interventions. Journal of Educational Psychology 108: 397–404.

Steele CM, Aronson, J. 1995. Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology 69: 797–811.

Steele CM. 1997. A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist 52: 613–629.

Stephens NM, Hamedani MG, Destin M. 2014. Closing the social-class achievement gap: A difference-education intervention improves first-generation students' academic performance and all students' college transition. Psychological Science 25: 943–953.

Taylor VJ, Walton GM. 2011. Stereotype threat undermines academic learning. Personality and Social Psychology Bulletin 37: 1055–1067.

Van Bavel JJ, Mende-Siedlecki P, Brady WJ, Reinero DA. 2016. Contextual sensitivity in scientific reproducibility. Proceedings of the National Academy of Sciences 113: 6454–6459.

Walton GM. 2014. The new science of wise psychological interventions. Current Directions in Psychological Science 23: 73–82.

Walton GM, Brady ST. 2020. The social-belonging intervention. Pages 36–62 in Walton GM, Crum AJ, eds. Handbook of Wise Interventions: How Social Psychology Can Help People Change. Guilford Press.

Walton GM, Cohen GL. 2007. A question of belonging: Race, social fit, and achievement. Journal of Personality and Social Psychology 92: 82–96.

Walton GM, Cohen GL. 2011. A brief social-belonging intervention improves academic and health outcomes of minority students. Science 331: 1447–1451.

Walton GM, Yeager DS. 2020. Seed and soil: Psychological affordances in contexts help to explain where wise interventions succeed or fail. Current Directions in Psychological Science 29: 219–226.

Walton GM, Logel C, Peach JM, Spencer SJ, Zanna MP. 2015. Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering. Journal of Educational Psychology 107: 468–485.

Widmeyer WN, Loy JW. 1988. When you're hot, you're hot! Warm–cold effects in first impressions of persons and teaching effectiveness. Journal of Educational Psychology 80: 118–121.

Wheeler CS, Petty RE. 2001.The effects of stereotype activation on behavior: A review of possible mechanisms. Psychological Bulletin 127: 797–826.

Yeager DS, et al. 2016. Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. Journal of Educational Psychology 108: 374–391.

Yeager DS, et al. 2021. Teacher mindsets help explain where a growth mindset intervention does and doesn't work. Psychological Science (22 December 2021): 1177/09567976211028984. https://doi.org/10.1177/09567976211028984.

Sarah Hammarlund (hamma111@umn.edu) is affiliated with the Department of Ecology, Evolution, and Behavior and with the BioTechnology Institute and Cheryl Scott (scot0136@umn.edu) is affiliated with the Department of Biology Teaching and Learning at the University of Minnesota, in Minneapolis, Minnesota, in the United States. Kevin Binning (kbinning@pitt.edu) is affiliated with the Department of Psychology and the Learning Research and Development Center at the University of Pittsburgh, in Pittsburgh, Pennsylvania, in the United States. Sehoya Cotner (sehoya@umn.edu) is affiliated with the Department of Biology Teaching and Learning at the University of Minnesota, in Minneapolis, Minnesota, in the United States, and with the Department of Biological Sciences at the University of Bergen, in Bergen, Norway.